2018 IEEE Conference on Decision and Control (CDC)
Miami Beach, FL, USA, Dec. 17-19, 2018

On the Location of the Minimizer of the Sum of Two Strongly Convex Functions

Kananart Kuwaranancharoen and Shreyas Sundaram

***Abstract*— The problem of finding the minimizer of a sum of convex functions is central to the field of distributed optimization. Thus, it is of interest to understand how that minimizer is related to the properties of the individual functions in the sum. In this paper, we provide an upper bound on the region containing the minimizer of the sum of two strongly convex functions. We consider two scenarios with different constraints on the upper bound of the gradients of the functions. In the first scenario, the gradient constraint is imposed on the location of the potential minimizer, while in the second scenario, the gradient constraint is imposed on a given convex set in which the minimizers of two original functions are embedded. We characterize the boundaries of the regions containing the minimizer in both scenarios.**

I. Introduction

The problem of distributed optimization arises in a variety of applications, including machine learning [1]–[4], control of large-scale systems [5], [6], and cooperative robotic systems [7]–[11]. In such problems, each node in a network has access to a local convex function (e.g., representing certain data available at that node), and all nodes are required to calculate the minimizer of the sum of the local functions. There is a significant literature on distributed algorithms that allow the nodes to achieve this objective [12]–[18]. The local functions in the above settings are typically assumed to be private to the nodes. However, there are certain common assumptions that are made about the characteristics of such functions, including strong convexity and bounds on the gradients (e.g., due to minimization over a convex set).

In certain applications, it may be of interest to determine a region where the minimizer of the sum of the functions can be located, given only the minimizers of the local functions, their strong convexity parameters, and the bound on their gradients (either at the minimizer or at the boundaries of a convex constraint set). For example, when the network contains malicious nodes that do not follow the distributed optimization algorithm, one cannot guarantee that all nodes calculate the true minimizer. Instead, one must settle for algorithms that allow the non-malicious nodes to converge to a certain region [19], [20]. In such situations, knowing the region where the minimizer can lie would allow us to evaluate the efficacy of such resilient distributed optimization algorithms. Similarly, suppose that the true functions at some (or all) nodes are not known (e.g., due to noisy data, or if the nodes obfuscate their functions due to privacy concerns). A key question in such scenarios is to determine how far the minimizer of the sum of the true functions can be from the minimizer calculated from the noisy (or obfuscated) functions. The region containing all possible minimizers of the sum of functions (calculated using only their local minimizers, convexity parameters, and bound on the gradients) would provide the answer to this question.

When the local functions f_i at each node v_i are single dimensional (i.e., $f_i : \mathbb{R} \to \mathbb{R}$), and strongly convex, it is easy to see that the minimizer of the sum of functions must be in the interval bracketed by the smallest and largest minimizers of the local functions. This is because the gradients of all the functions will have the same sign outside that region, and thus cannot sum to zero. However, a similar characterization of the region containing the minimizer of multidimensional functions is lacking in the literature, and is significantly more challenging to obtain. For example, the conjecture that the minimizer of a sum of convex functions is in the convex hull of their local minimizers can be easily disproved via simple examples; consider $f_1(x, y) = x^2 - xy + \frac{1}{2}y^2$ and $f_2(x, y) = x^2 + xy + \frac{1}{2}y^2 - 4x - 2y$ with minimizers $(0, 0)$ and $(2, 0)$ respectively, whose sum has minimizer $(1, 1)$. Thus, in this paper, **our goal is to take a step toward characterizing the region containing the minimizer of a sum of strongly convex functions.** Specifically, we focus on characterizing this region for the sum of *two* strongly convex functions under various assumptions on their gradients (as described in the next section). As we will see, the analysis is significantly complicated even for this scenario. Nevertheless, we obtain such a region and gain insights that could potentially be leveraged in future work to tackle the sum of multiple functions.

II. Notation and Preliminaries

Sets: We denote the closure and interior of a set $\mathcal{E}$ by $\bar{\mathcal{E}}$ and $\mathcal{E}^\circ$, respectively. The boundary of a set $\mathcal{E}$ defined as $\partial\mathcal{E} = \bar{\mathcal{E}} \setminus \mathcal{E}^\circ$.

Linear Algebra: We denote by $\mathbb{R}^n$ the n-dimensional Euclidean space. For simplicity, we often use $(x_1, \ldots, x_n)$ to represent the column vector $\begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}^T$. We use e_i to denote the i-th basis vector (the vector of all zeros except for a one in the i-th position). We denote by $\|\cdot\|$ the Euclidean norm $\|x\| := (\sum_i x_i^2)^{1/2}$ and by $\angle(u, v)$ the angle between vectors u and v. Note that $\angle(u, v) = \arccos\big(\frac{u^T v}{\|u\|\|v\|}\big)$. We use $\mathcal{B}_r(x_0) = \{x \in \mathbb{R}^n : \|x - x_0\| < r\}$ and $\bar{\mathcal{B}}_r(x_0)$ to denote the open and closed ball, respectively, centered at x_0 of radius r.

Convex Sets and Functions: A set $\mathcal{C}$ in $\mathbb{R}^n$ is said to be convex if, for all x and y in $\mathcal{C}$ and all t in the interval $(0, 1)$,

This research was supported by NSF CAREER award 1653648. The authors are with the School of Electrical and Computer Engineering at Purdue University. Email: {kkuwaran, sundara2}@purdue.edu.

the point $(1-t)x + ty$ also belongs to $\mathcal{C}$. A differentiable function f is called strongly convex with parameter $\sigma > 0$ (or σ-strongly convex) if $(\nabla f(x) - \nabla f(y))^T(x-y) \geq \sigma\|x-y\|^2$ holds for all points x, y in its domain. We denote the set of all σ-strongly convex functions by $\mathcal{S}(\sigma)$.

III. Problem Statement

We will consider two scenarios in this paper. We first consider constraints on the gradients of the local functions at the location of the potential minimizer, and then consider constraints on the gradients inside a convex constraint set.

A. Problem 1

Consider two strongly convex functions $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$. The two functions f_1 and f_2 have strong convexity parameters σ_1 and σ_2, respectively, and minimizers x_1^* and x_2^*, respectively. Let x denote the minimizer of $f_1 + f_2$, and suppose that the norm of the gradients of f_1 and f_2 must be bounded above by a finite number L at x. Our goal is to estimate the region $\mathcal{M}$ containing all possible values x satisfying the above conditions. More specifically, we wish to estimate the region

$$\begin{aligned}\mathcal{M}(x_1^*, x_2^*, \sigma_1, \sigma_2, L) \triangleq \{x \in \mathbb{R}^n : \exists f_1 \in \mathcal{S}(\sigma_1),\\ \exists f_2 \in \mathcal{S}(\sigma_2),\ \ \nabla f_1(x_1^*) = 0,\ \ \nabla f_2(x_2^*) = 0,\\ \nabla f_1(x) = -\nabla f_2(x),\ \ \|\nabla f_1(x)\| = \|\nabla f_2(x)\| \leq L\}. \quad (1)\end{aligned}$$

For simplicity of notation, we will omit the argument of the set $\mathcal{M}(x_1^*, x_2^*, \sigma_1, \sigma_2, L)$ and write it as $\mathcal{M}$ or $\mathcal{M}(x_1^*, x_2^*)$.

B. Problem 2

Consider two strongly convex functions $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$. The two functions f_1 and f_2 have strong convexity parameters σ_1 and σ_2, respectively, and minimizers x_1^* and x_2^*, respectively. Suppose that we also have a compact convex set $\mathcal{C} \subset \mathbb{R}^n$ containing the minimizers x_1^* and x_2^*. Let x denote the minimizer of $f_1 + f_2$ within the region $\mathcal{C}$. The norm of the gradients of both functions f_1 and f_2 is bounded above by a finite number L everywhere in the set $\mathcal{C}$. Our goal is to estimate the region $\mathcal{N}$ containing all possible values $x_0 \in \mathcal{C}$ satisfying the above conditions. More specifically, define $\mathcal{F}(\sigma, L, \mathcal{C})$ to be the family of functions that are σ-strongly convex and whose gradient norm is upper bounded by L everywhere inside the convex set $\mathcal{C}$:

$$\mathcal{F}(\sigma, L, \mathcal{C}) \triangleq \{f : f \in \mathcal{S}(\sigma),\ \|\nabla f(x)\| \leq L,\ \forall x \in \mathcal{C}\}.$$

Then, we wish to characterize the region

$$\begin{aligned}\mathcal{N}(x_1^*, x_2^*, \sigma_1, \sigma_2, L) \triangleq \{x \in \mathbb{R}^n : \exists f_1 \in \mathcal{F}(\sigma_1, L, \mathcal{C}),\\ \exists f_2 \in \mathcal{F}(\sigma_2, L, \mathcal{C}),\quad \nabla f_1(x_1^*) = 0,\\ \nabla f_2(x_2^*) = 0,\quad \nabla f_1(x) = -\nabla f_2(x)\}. \quad (2)\end{aligned}$$

For simplicity of notation, we will omit the argument of the set $\mathcal{N}(x_1^*, x_2^*, \sigma_1, \sigma_2, L)$ and write it as $\mathcal{N}$ or $\mathcal{N}(x_1^*, x_2^*)$.

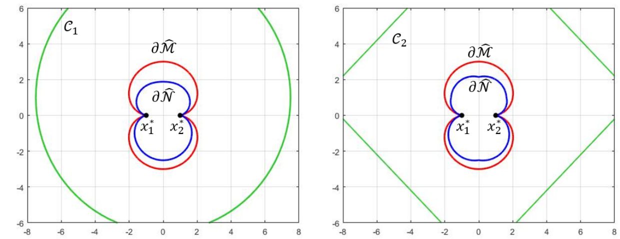

Fig. 1. The red lines are the boundary of the region that contains $\mathcal{M}$, while the blue lines are the boundary of the region that contains $\mathcal{N}$, where convex sets $\mathcal{C}_1$ and $\mathcal{C}_2$ are a circle (Left) and a box (Right) respectively.

C. A Preview of the Solution

We provide two examples of the region containing the minimizer of the sum of 2-dimensional functions in both scenarios in Fig. 1, where x_1^* and x_2^* are the minimizers of f_1 and f_2, respectively; we derive these regions in the rest of the paper. Notice that the region containing set $\mathcal{M}$ (the area bounded by the red line) is bigger than the region containing set $\mathcal{N}$ (the area bounded by the blue line). In addition, even though we have changed the shape of convex set in the two examples, the minimizer regions are similar.

IV. Problem 1: Gradient Constraint at Location of Potential Minimizer

In this section, we consider the first scenario when the gradient constraint is imposed on the location of the potential minimizer and derive an approximation to the set $\mathcal{M}$ in (1).

Consider functions $f_1 \in \mathcal{S}(\sigma_1)$ with minimizer x_1^* and $f_2 \in \mathcal{S}(\sigma_2)$ with minimizer x_2^*. Without loss of generality, we can assume $x_1^* = (-r, 0, \ldots, 0) \in \mathbb{R}^n$ and $x_2^* = (r, 0, \ldots, 0) \in \mathbb{R}^n$ for some $r \in \mathbb{R}_{>0}$, since for any x_1^* and x_2^* such that $x_1^* \neq x_2^*$, we can find a unique affine transformation that maps the original minimizers into these values and also preserves the distance between these points i.e., $\|x_1^* - x_2^*\| = 2r$. The minimizer region in the original coordinates can then be obtained by applying the inverse transformation to the derived region.

We will be using the following functions throughout our analysis. For $i \in \{1, 2\}$, define

$$\tilde{\phi}_i(x, L) \triangleq \arccos\left(\frac{\sigma_i}{L}\|x - x_i^*\|\right), \quad (3)$$

for all $x \in \mathbb{R}^n$ such that $\frac{\sigma_i}{L}\|x - x_i^*\| \leq 1$. For simplicity of notation, if L is a constant, we will omit the arguments and write it as $\tilde{\phi}_i(x)$ or $\tilde{\phi}_i$. Furthermore, for all $x \in \mathbb{R}^n$, define

$$\psi(x) \triangleq \pi - (\alpha_2(x) - \alpha_1(x)),$$

where $\alpha_i(x)$ is the angle between $x - x_i^*$ and $x_2^* - x_1^*$ i.e., $\alpha_i(x) \triangleq \angle(x - x_i^*, x_2^* - x_1^*)$.

Lemma 1: Necessary conditions for a point $x \in \mathbb{R}^n$ to be a minimizer of $f_1 + f_2$ when the gradients of f_1 and f_2 are bounded by L at x are (i) $\|x - x_i^*\| \leq \frac{L}{\sigma_i}$ for $i = 1, 2$, and (ii) $\tilde{\phi}_1(x) + \tilde{\phi}_2(x) \geq \psi(x)$.

Proof: From the definition of strongly convex functions,

$$(\nabla f_i(x) - \nabla f_i(y))^T(x - y) \geq \sigma_i\|x - y\|^2$$

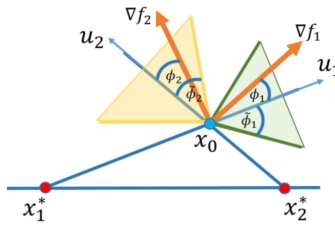

Fig. 2. The quantities $\phi_i(x_0)$ represent the angles between $\nabla f_i(x_0)$ and $u_i(x_0)$. The quantities $\tilde{\phi}_i(x_0)$ represent the maximum possible values for $\phi_i(x_0)$ in order for x_0 to be a minimizer. In other words, the angles $\phi_1(x_0)$ and $\phi_2(x_0)$ must lie in the shaded regions.

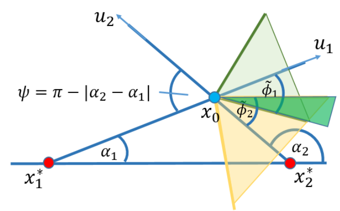

Fig. 3. The green region in the figure is the set $\mathcal{G}_1(x_0)$ and the yellow region is the set $\mathcal{G}_2(x_0)$. These regions are defined by the angles $\tilde{\phi}_1$ and $\tilde{\phi}_2$. If these regions overlap, the point x_0 is a minimizer candidate.

for all x, y and for $i = 1, 2$. Since x_1^* and x_2^* are the minimizers of f_1 and f_2 respectively, we get

$$(\nabla f_i(x) - \nabla f_i(x_i^*))^T(x - x_i^*) \geq \sigma_i \|x - x_i^*\|^2$$
$$\Rightarrow \quad \nabla f_i(x)^T \frac{x - x_i^*}{\|x - x_i^*\|} \geq \sigma_i \|x - x_i^*\| \geq 0. \quad (4)$$

Let $u_i(x) \triangleq \frac{x - x_i^*}{\|x - x_i^*\|}$ be the unit vector in the direction of $x - x_i^*$ and $\phi_i(x) \triangleq \angle(\nabla f_i(x), u_i(x))$, with $0 \leq \phi_i(x) \leq \frac{\pi}{2}$ as shown in Fig. 2. From (4), we get

$$\nabla f_i(x)^T u_i(x) = \|\nabla f_i(x)\| \cos(\phi_i(x)) \geq \sigma_i \|x - x_i^*\|.$$

If x is a candidate minimizer then we can apply the gradient norm constraint $\|\nabla f_i(x)\| \leq L$ to the above inequality to obtain

$$\cos(\phi_i(x)) \geq \frac{\sigma_i}{L} \|x - x_i^*\|. \quad (5)$$

If $\frac{\sigma_i}{L}\|x - x_i^*\| \leq 1$ then $\phi_i(x) \leq \arccos(\frac{\sigma_i}{L}\|x - x_i^*\|)$. On the other hand, if $\frac{\sigma_i}{L}\|x - x_i^*\| > 1$ then there is no $\phi_i(x)$ that can satisfy the inequality (5). Therefore, if $\|x - x_1^*\| > \frac{L}{\sigma_1}$ or $\|x - x_2^*\| > \frac{L}{\sigma_2}$, we conclude that x cannot be the minimizer of the function $f_1 + f_2$.

Suppose that $\frac{\sigma_i}{L}\|x - x_i^*\| \leq 1$ for $i = 1, 2$ so that $\arccos(\frac{\sigma_1}{L}\|x - x_1^*\|)$ and $\arccos(\frac{\sigma_2}{L}\|x - x_2^*\|)$ are well-defined. In order to capture the possible gradient of f_1 at point x, define a set of vectors whose norms are at most L and satisfy (5):

$$\mathcal{G}_1(x) \triangleq \Big\{g : \|g\| \leq L, \\ \angle(g, u_1(x)) \leq \arccos\left(\frac{\sigma_1}{L}\|x - x_1^*\|\right)\Big\}.$$

Since x can be the minimizer of the function $f_1 + f_2$ only when $\nabla f_1(x) = -\nabla f_2(x)$, we define a set of vectors whose norms are at most L and satisfy (5) to capture the possible negated gradient vectors of f_2:

$$\mathcal{G}_2(x) \triangleq \Big\{g : \|g\| \leq L, \\ \angle(-g, u_2(x)) \leq \arccos\left(\frac{\sigma_2}{L}\|x - x_2^*\|\right)\Big\}.$$

Note that $\phi_2(x)$ can be viewed geometrically as the angle between $-\nabla f_2(x)$ and $-u_2(x)$ as shown in Fig. 2. If $\mathcal{G}_1(x) \cap \mathcal{G}_2(x) = \emptyset$, then x cannot be the minimizer of the function $f_1 + f_2$ because it is not possible to choose f_1 and f_2 such that $\nabla f_i(x)$ satisfy inequality (5) for $i = 1, 2$ and $\nabla f_1(x) = -\nabla f_2(x)$ simultaneously.

Recall that $\alpha_i(x) = \angle(u_i(x), x_2^* - x_1^*)$ with $0 \leq \alpha_i(x) \leq \pi$ for $i = 1, 2$, i.e., $\alpha_i(x) = \arccos\left(u_i(x)^T \frac{(x_2^* - x_1^*)}{\|x_2^* - x_1^*\|}\right)$. Note that $\alpha_2(x) \geq \alpha_1(x)$ due to the definition of α_i. Then, the angle between $u_1(x)$ and $u_2(x)$ is $\alpha_2(x) - \alpha_1(x)$. Therefore, the angle between $u_1(x)$ and $-u_2(x)$ is equal to $\psi(x) = \pi - (\alpha_2(x) - \alpha_1(x))$.

Let $\tilde{\phi}_i(x)$ be the maximum angle of $\phi_i(x)$ that satisfies inequality (5), i.e., as given by (3). By the definition of $\tilde{\phi}_i(x)$, if $\tilde{\phi}_1(x) + \tilde{\phi}_2(x) \geq \psi(x)$, there is an overlapping region caused by $\tilde{\phi}_1(x)$ and $\tilde{\phi}_2(x)$ as shown in Fig. 3 and there exist gradients $\nabla f_1(x) \in \mathcal{G}_1(x)$ and $-\nabla f_2(x) \in \mathcal{G}_2(x)$ such that $\nabla f_1(x) = -\nabla f_2(x)$. On the other hand, if $\tilde{\phi}_1(x) + \tilde{\phi}_2(x) < \psi(x)$ then $\mathcal{G}_1(x) \cap \mathcal{G}_2(x) = \emptyset$ and it is not possible to choose gradients $\nabla f_1(x) \in \mathcal{G}_1(x)$ and $-\nabla f_2(x) \in \mathcal{G}_2(x)$ such that they cancel each other. In this case, we can conclude that this x cannot be the minimizer of the function $f_1 + f_2$. ■

Note that angles $\tilde{\phi}_1(x)$, $\tilde{\phi}_2(x)$, $\alpha_1(x)$, and $\alpha_2(x)$ can be expressed as a function of $\|x_1^* - x_2^*\|$, $\|x - x_1^*\|$, and $\|x - x_2^*\|$. Thus, from the proof of Lemma 1, the inequality $\tilde{\phi}_1(x) + \tilde{\phi}_2(x) \geq \psi(x)$ depends only on the distance between the three points x_1^*, x_2^*, and x. Therefore, the candidate minimizer property of x can be fully described by the 2-D picture in Fig. 3.

Now we consider the relationship between set $\mathcal{M}$ in (1) (which is the set that we want to identify) and certain other sets which we define below. Define the set

$$\hat{\mathcal{M}}(x_1^*, x_2^*) \triangleq \Big\{x \in \mathbb{R}^n : \tilde{\phi}_1(x) + \tilde{\phi}_2(x) \geq \psi(x), \\ \|x - x_1^*\| \leq \frac{L}{\sigma_1}, \quad \|x - x_2^*\| \leq \frac{L}{\sigma_2}\Big\}.$$

Note that based on Lemma 1, $\hat{\mathcal{M}}$ contains the minimizers of $f_1 + f_2$.

Define $\mathcal{H}$ to be the set of points such that there exist strongly convex functions (with given strong convexity parameters and minimizers) whose gradients can be bounded by L at those points:

$$\mathcal{H}(x_1^*, x_2^*) \triangleq \{x \in \mathbb{R}^n : \exists f_1 \in \mathcal{S}(\sigma_1), \\ \exists f_2 \in \mathcal{S}(\sigma_2), \quad \nabla f_1(x_1^*) = 0, \quad \nabla f_2(x_2^*) = 0, \\ \|\nabla f_1(x)\| \leq L, \quad \|\nabla f_2(x)\| \leq L\}.$$

Define $\mathcal{H}_i$ to be the set of points such that there exists a σ_i-strongly convex function f_i with minimizer x_i^* whose gradient is bounded by L at those points:

$$\mathcal{H}_i(x_i^*) \triangleq \{x \in \mathbb{R}^n : \exists f_i \in \mathcal{S}(\sigma_i), \nabla f_i(x_i^*) = 0,\ \|\nabla f_i(x)\| \le L\},\ \ i = 1, 2.$$

Lemma 2: $\mathcal{M}(x_1^*, x_2^*) \subseteq \hat{\mathcal{M}}(x_1^*, x_2^*) \subseteq \mathcal{H}(x_1^*, x_2^*)$ and $\mathcal{H}(x_1^*, x_2^*) = \bar{\mathcal{B}}_{\frac{L}{\sigma_1}}(x_1^*) \cap \bar{\mathcal{B}}_{\frac{L}{\sigma_2}}(x_2^*)$.

Proof: From Lemma 1, we get $\mathcal{M}(x_1^*, x_2^*) \subseteq \hat{\mathcal{M}}(x_1^*, x_2^*)$. From the definition of a strongly convex function,

$$(\nabla f_i(x) - \nabla f_i(y))^T (x - y) \ge \sigma_i \|x - y\|^2$$

for all x, y where $i = 1, 2$. Substitute x_i^* into y to get

$$\begin{aligned} (\nabla f_i(x) - \nabla f_i(x_i^*))^T (x - x_i^*) &\ge \sigma_i \|x - x_i^*\|^2 \\ \Leftrightarrow \|\nabla f_i(x)\| \|x - x_i^*\| \cos(\phi_i(x)) &\ge \sigma_i \|x - x_i^*\|^2 \\ \Rightarrow \quad L &\ge \sigma_i \|x - x_i^*\| \\ \Leftrightarrow \quad \|x - x_i^*\| &\le \frac{L}{\sigma_i} \end{aligned} \tag{6}$$

where the equality $\|\nabla f_i(x)\| \cos(\phi_i(x)) = L$ occurs when $\nabla f_i(x)$ is chosen such that $\|\nabla f_i(x)\| = L$ and $\nabla f_i(x)^T u_i(x) = L$. Note that the above sequence of inequalities uses the fact that $\|\nabla f_i(x)\| \le L$ and $0 \le \cos(\phi_i(x)) \le 1$. Since $\bar{\mathcal{B}}_r(x_0) = \{x : \|x - x_0\| \le r\}$, from (6), we have $\mathcal{H}_i(x_i^*) \subseteq \bar{\mathcal{B}}_{\frac{L}{\sigma_i}}(x_i^*)$.

For the converse, consider $\hat{x} \in \bar{\mathcal{B}}_{\frac{L}{\sigma_i}}(x_i^*)$. By choosing a quadratic function $f_i(x) = \frac{1}{2}\hat{\sigma}_i (x - x_i^*)^T (x - x_i^*)$ where $\hat{\sigma}_i = \frac{L}{\|\hat{x} - x_i^*\|}$, one can easily verify that $\hat{\sigma}_i \ge \sigma_i$ and $\|\nabla f_i(\hat{x})\| = L$. So, we have $\mathcal{H}_i(x_i^*) \supseteq \bar{\mathcal{B}}_{\frac{L}{\sigma_i}}(x_i^*)$.

From the definition of $\mathcal{H}$ and $\mathcal{H}_i$, we get $\mathcal{H}(x_1^*, x_2^*) = \mathcal{H}_1(x_1^*) \cap \mathcal{H}_2(x_2^*) = \bar{\mathcal{B}}_{\frac{L}{\sigma_1}}(x_1^*) \cap \bar{\mathcal{B}}_{\frac{L}{\sigma_2}}(x_2^*)$. Finally, since the conditions of the set $\mathcal{H}$ are the same as the last two conditions in the set $\hat{\mathcal{M}}$, we get $\hat{\mathcal{M}}(x_1^*, x_2^*) \subseteq \mathcal{H}(x_1^*, x_2^*)$. ∎

The result from Lemma 2 shows that the set $\hat{\mathcal{M}}$ contains the set $\mathcal{M}$ from (1) within it. Thus, we will derive the equation of the boundary of $\hat{\mathcal{M}}$ in n-dimensional space from the angles $\tilde{\phi}_i$ defined in (3), and the necessary condition $\tilde{\phi}_1(x) + \tilde{\phi}_2(x) \ge \psi(x)$.

Define $x = (z_1, z_2, \ldots, z_n)$ and the set of points

$$\mathcal{T}_n(L) \triangleq \left\{ x \in \mathbb{R}^n : \frac{z_1^2 + \|\mathbf{z}\|^2 - r^2}{d_1^2 d_2^2} + \frac{\sigma_1 \sigma_2}{L^2} = \sqrt{\frac{1}{d_1^2} - \frac{\sigma_1^2}{L^2}} \cdot \sqrt{\frac{1}{d_2^2} - \frac{\sigma_2^2}{L^2}} \right\}$$

where $d_1(x) = \sqrt{(z_1 + r)^2 + \|\mathbf{z}\|^2}$, $d_2(x) = \sqrt{(z_1 - r)^2 + \|\mathbf{z}\|^2}$, $z_1 \in \mathbb{R}$ and $\mathbf{z} = (z_2, z_3, \ldots, z_n) \in \mathbb{R}^{n-1}$. For simplicity of notation, if L is a constant, we will omit the argument and write it as $\mathcal{T}_n$.

Lemma 3: The set $\{x : \tilde{\phi}_1(x) + \tilde{\phi}_2(x) = \pi - (\alpha_2(x) - \alpha_1(x))\}$ is equivalent to $\mathcal{T}_n$.

Proof: From Fig. 3, the z_1-axis equations are given by (with x elided for notational convenience)

$$\begin{aligned} z_1 &= d_1 \cos\alpha_1 - r = d_2 \cos\alpha_2 + r, \\ \Leftrightarrow \quad \cos\alpha_1 &= \frac{z_1 + r}{d_1} \quad \text{and} \quad \cos\alpha_2 = \frac{z_1 - r}{d_2}. \end{aligned} \tag{7}$$

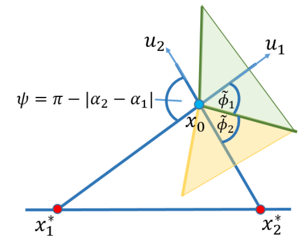

Fig. 4. The sets of gradients at a point on the boundary $\partial\hat{\mathcal{M}}$ that is not on the boundary $\partial\mathcal{H}$. In this case, $\tilde{\phi}_1(x_0) + \tilde{\phi}_2(x_0) = \psi(x_0)$.

The $\mathbf{z}$-axes equations are given by

$$\begin{aligned} \|\mathbf{z}\| &= d_1 \sin\alpha_1 = d_2 \sin\alpha_2, \\ \Leftrightarrow \quad \sin\alpha_1 &= \frac{\|\mathbf{z}\|}{d_1} \quad \text{and} \quad \sin\alpha_2 = \frac{\|\mathbf{z}\|}{d_2}. \end{aligned} \tag{8}$$

Consider

$$\tilde{\phi}_1(x) + \tilde{\phi}_2(x) = \pi - (\alpha_2(x) - \alpha_1(x)). \tag{9}$$

Since $0 \le \tilde{\phi}_i \le \frac{\pi}{2}$, we get $0 \le \tilde{\phi}_1 + \tilde{\phi}_2 \le \pi$. Since $0 \le \alpha_i \le \pi$ and $\alpha_2 \ge \alpha_1$, $0 \le \pi - (\alpha_2 - \alpha_1) \le \pi$. Thus, equation (9) is equivalent to

$$\begin{aligned} &\cos(\tilde{\phi}_1 + \tilde{\phi}_2) = \cos(\pi - (\alpha_2 - \alpha_1)) \\ \Leftrightarrow \quad &\cos(\tilde{\phi}_1 + \tilde{\phi}_2) = -\cos(\alpha_2 - \alpha_1). \end{aligned}$$

Expanding this equation and substituting (7), (8), and $\cos(\tilde{\phi}_i(x)) = \frac{\sigma_i}{L} d_i$ for $i = 1, 2$, we get

$$\frac{\sigma_1}{L} d_1 \cdot \frac{\sigma_2}{L} d_2 - \sqrt{1 - \left(\frac{\sigma_1}{L} d_1\right)^2} \cdot \sqrt{1 - \left(\frac{\sigma_2}{L} d_2\right)^2} = -\frac{z_1 - r}{d_2} \cdot \frac{z_1 + r}{d_1} - \frac{\|\mathbf{z}\|}{d_2} \cdot \frac{\|\mathbf{z}\|}{d_1}.$$

Dividing the above equation by $d_1 d_2$ and rearranging yields $\mathcal{T}_n$. ∎

We next provide a lemma that will subsequently lead to the main result of this section, namely Theorem 1. The proof of Lemma 4 and Theorem 1 can be found in [21].

Lemma 4: (i) $x_1^* \in \partial\hat{\mathcal{M}}$ if and only if $r \le \frac{L}{2\sigma_2}$.
(ii) $x_2^* \in \partial\hat{\mathcal{M}}$ if and only if $r \le \frac{L}{2\sigma_1}$.

Theorem 1: If $r \le \frac{L}{2} \cdot \min\left\{\frac{1}{\sigma_1}, \frac{1}{\sigma_2}\right\}$ then the boundary $\partial\hat{\mathcal{M}}$ is given by $\mathcal{T}_n \cup \{x_1^*, x_2^*\}$.

First, we provide a sketch of the proof of Theorem 1 here. Using Lemma 4, we can conclude that x_1^* and x_2^* are included in the boundary $\partial\hat{\mathcal{M}}$ under the conditions of the theorem.

Then, consider the case when $x \notin \{x_1^*, x_2^*\}$. Recall the definition of $\hat{\mathcal{M}}$ and Lemma 2. The boundary $\partial\hat{\mathcal{M}}$ can be classified into 2 disjoint types. The first type consists of points x with the following property: $\tilde{\phi}_1(x) + \tilde{\phi}_2(x) = \pi - (\alpha_2(x) - \alpha_1(x))$ for which an example is shown in Fig. 4. The second type consists of points x with the following property: $\tilde{\phi}_1(x) + \tilde{\phi}_2(x) > \pi - (\alpha_2(x) - \alpha_1(x))$ for which an example is shown in Fig. 5. One can show that if x is in the set $\{x : \tilde{\phi}_1(x) + \tilde{\phi}_2(x) > \psi(x)\}$, then $x \notin \partial\hat{\mathcal{M}}$. Therefore,

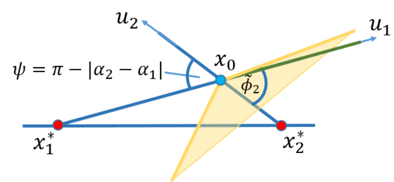

Fig. 5. The sets of gradients at a point on the boundary $\partial\hat{\mathcal{M}}$ that is also on the boundary $\partial\mathcal{H}$. In this case, $\cos(\tilde{\phi}_1(x_0)) = 1$ and $\cos(\tilde{\phi}_2(x_0)) = \frac{\sigma_2}{L}\|x_0 - x_2^*\|$; however, $\tilde{\phi}_1(x_0) + \tilde{\phi}_2(x_0) > \psi(x_0)$.

$\partial\hat{\mathcal{M}}$ must be described by the set $\{x : \tilde{\phi}_1(x) + \tilde{\phi}_2(x) = \psi(x)\}$ (which is $\mathcal{T}_n$ by Lemma 3), along with $\{x_1^*, x_2^*\}$.

An example of the region $\hat{\mathcal{M}}$ given by Theorem 1 is shown in Fig. 6.

V. Problem 2: Gradient Constraint on Convex Set

In this section, we consider the second scenario when the gradient constraint is imposed on a given convex set in which the minimizers of two original functions are embedded. We begin by analyzing the necessary condition for any given point to be a minimizer using a geometric approach and then state the relationship among certain sets related to the minimizer region. Finally, the equation of a region of possible minimizers in n-dimensional space is presented. For Lemma 5 and 6, and Theorem 2, we will discuss the main ideas of the proof briefly. The complete proof is provided at [21].

Let $d(x_0, \partial\mathcal{C})$ be the infimum distance between x_0 and the boundary of a convex set $\mathcal{C}$, i.e.,

$$d(x_0, \partial\mathcal{C}) \triangleq \inf_{x\in\partial\mathcal{C}} \|x - x_0\|.$$

Lemma 5: Suppose $\mathcal{C}$ is a compact convex set and x_0 is a point in $\mathcal{C}$. Suppose $f \in \mathcal{S}(\sigma)$, and the norm of the gradient of f in $\mathcal{C}$ is bounded by L, i.e., $\|\nabla f(x)\| \le L$, $\forall x \in \mathcal{C}$. Then

$$\|\nabla f(x_0)\| \le L - \sigma d(x_0, \partial\mathcal{C})$$

The main idea of the proof of this lemma is that the norm of the gradient at x_0 (i.e., $\|\nabla f(x_0)\|$) plus the additional gradient increase from x_0 to the boundary $\partial\mathcal{C}$ must not exceed L. However, the distance from x_0 to the boundary is bounded below by $d(x_0, \partial C)$, so by rearranging the inequality, we obtain the result.

Lemma 6: Suppose $\mathcal{C}$ is a compact convex set. Let $f_1 \in \mathcal{S}(\sigma_1)$, $f_2 \in \mathcal{S}(\sigma_2)$, x_0 be the minimizer of $f_1 + f_2$ over the set $\mathcal{C}$ and $\hat{L}$ be the norm of the gradient of f_1 and f_2 at x_0. If the norm of the gradient of f_1 and f_2 in $\mathcal{C}$ is bounded by L, i.e., $\|\nabla f_i(x)\| \le L$, $\forall x \in \mathcal{C}$, $i = 1, 2$ then

$$\hat{L} \le L - \min(\sigma_1, \sigma_2) \times d(x_0, \partial\mathcal{C}).$$

In order to prove this lemma, we use the result from Lemma 5 and apply it to the functions f_1 and f_2. Since the gradients at x_0 are equal i.e., $\|\nabla f_1(x_0)\| = \|\nabla f_2(x_0)\|$, the minimum growth rate of gradient from x_0 to x is determined by $\min(\sigma_1, \sigma_2)$.

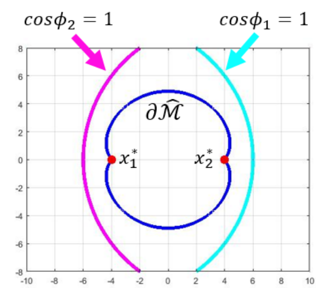

Fig. 6. The boundary $\partial\hat{\mathcal{M}}$ (blue line) is plotted given original minimizers $x_1^* = (-4, 0)$ and $x_2^* = (4, 0)$ and parameters $\sigma_1 = \sigma_2 = 1$ and $L = 10$.

As before, without loss of generality, we can assume $x_1^* = (-r, 0, \ldots, 0) \in \mathbb{R}^n$ and $x_2^* = (r, 0, \ldots, 0) \in \mathbb{R}^n$ since for any minimizers x_1^* and x_2^*, and a convex set $\mathcal{C}$, we can find a unique affine transformation that maps the original minimizers into $(-r, 0, \ldots, 0)$ and $(r, 0, \ldots, 0)$ respectively and also preserves the distance between these points, i.e., $\|x_1^* - x_2^*\| = 2r$. This transformation also uniquely maps the original convex set $\mathcal{C}$ into a new convex set $\mathcal{C}'$.

With the above assumption, we can now modify Lemma 1 with the new bound $\hat{L}$ on $\|\nabla f_i(x_0)\|$, provided by Lemma 6. Define a function

$$\tilde{L}(x) \triangleq L - \min(\sigma_1, \sigma_2) \times d(x, \partial\mathcal{C}) \quad \text{for} \quad x \in \mathcal{C}.$$

Lemma 7: Necessary conditions for a point $x \in \mathbb{R}^n$ to be a minimizer of $f_1 + f_2$ when the gradients of f_1 and f_2 are bounded by L in a convex set $\mathcal{C}$ are (i) $\|x - x_i^*\| \le \frac{1}{\sigma_i}\tilde{L}(x)$ for $i = 1, 2$, and (ii) $\tilde{\phi}_1(x, \tilde{L}) + \tilde{\phi}_2(x, \tilde{L}) \ge \psi(x)$.

The proof is the same as Lemma 1 except that we use $\|\nabla f_i(x)\| \le \tilde{L}(x)$ instead of $\|\nabla f_i(x)\| \le L$.

Now we consider the relationship between the set $\mathcal{N}$ in (2) (which is the set that we want to identify) and other sets which we will define below. Recall the definition of $\mathcal{N}$ from (2) where $\mathcal{F}(\sigma, L, \mathcal{C}) = \{f : f \in \mathcal{S}(\sigma), \quad \|\nabla f(x)\| \le L, \quad \forall x \in \mathcal{C}\}$ for a given convex set $\mathcal{C}$.

We define $\hat{\mathcal{N}}$ as

$$\hat{\mathcal{N}}(x_1^*, x_2^*) \triangleq \Big\{x \in \mathbb{R}^n : \tilde{\phi}_1(x, \tilde{L}) + \tilde{\phi}_2(x, \tilde{L}) \ge \psi(x), \\ \|x - x_1^*\| \le \frac{1}{\sigma_1}\tilde{L}(x), \quad \|x - x_2^*\| \le \frac{1}{\sigma_2}\tilde{L}(x)\Big\}$$

where $\tilde{L}(x) = L - \min(\sigma_1, \sigma_2) \times d(x, \partial\mathcal{C})$. Note that unlike L, $\tilde{L}(x)$ is a function of x. By Lemma 7, $\hat{\mathcal{N}}$ contains the minimizers of $f_1 + f_2$ and $\mathcal{N}(x_1^*, x_2^*) \subseteq \hat{\mathcal{N}}(x_1^*, x_2^*)$.

Define $\mathcal{I}$ to be the set

$$\mathcal{I}(x_1^*, x_2^*) \triangleq \{x \in \mathbb{R}^n : \exists f_1 \in \mathcal{S}(\sigma_1), \quad \exists f_2 \in \mathcal{S}(\sigma_2), \\ \nabla f_1(x_1^*) = 0, \quad \nabla f_2(x_2^*) = 0, \\ \|\nabla f_1(x)\| \le \tilde{L}(x), \quad \|\nabla f_2(x)\| \le \tilde{L}(x)\}.$$

Define $\mathcal{I}_i$, $i = 1, 2$, to be the set of points such that there exists a strongly convex function f_i whose minimizer is x_i^*

and whose gradient can be bounded by $\tilde{L}$ at x:

$$\mathcal{I}_i(x_i^*) \triangleq \{x \in \mathbb{R}^n : \exists f_i \in \mathcal{S}(\sigma_i),\ \nabla f_i(x_i^*) = 0, \\ \|\nabla f_i(x)\| \leq \tilde{L}(x)\}.$$

Lemma 8: $\mathcal{N}(x_1^*, x_2^*) \subseteq \hat{\mathcal{N}}(x_1^*, x_2^*) \subseteq \mathcal{I}(x_1^*, x_2^*)$, $\mathcal{I}(x_1^*, x_2^*) = \mathcal{I}_1(x_1^*) \cap \mathcal{I}_2(x_2^*)$, and $\hat{\mathcal{N}}(x_1^*, x_2^*) \subseteq \hat{\mathcal{M}}(x_1^*, x_2^*)$ for all $x \in \mathcal{C}$.

Proof: The first and second parts are similar to the proof of Lemma 2. However, we cannot simplify the set $\mathcal{I}_i$ further (unlike the set $\mathcal{H}_i$ in Lemma 2) since $\mathcal{I}_i$ depends on the convex set $\mathcal{C}$ (via $\tilde{L}$).

Since the gradient $\tilde{L}(x)$ is no greater than L for all $x \in \mathcal{C}$, the third part $\hat{\mathcal{N}}(x_1^*, x_2^*) \subseteq \hat{\mathcal{M}}(x_1^*, x_2^*)$ follows. ■

We can interpret Lemma 8 as follows. The constraints $\exists f_i \in \mathcal{F}(\sigma_i, L, \mathcal{C})$ for $i = 1, 2$ in the set $\mathcal{N}$ are shifted to a constraint on their gradients, i.e., $\tilde{L}(x) = L - \min(\sigma_1, \sigma_2) \times d(x, \partial\mathcal{C})$. This simplifies the analysis significantly but potentially introduces conservatism.

Theorem 2: If $\hat{\mathcal{M}}(x_1^*, x_2^*) \subset \mathcal{I}(x_1^*, x_2^*)$ and $r \leq \frac{L}{2} \times \min\left\{\frac{1}{\sigma_1}, \frac{1}{\sigma_2}\right\}$, then $\partial\hat{\mathcal{N}}$ is given by $\mathcal{T}_n(\tilde{L}) \cup \{x_1^*, x_2^*\}$.

The proof of Theorem 2 in [21] is obtained by noting from Theorem 1 and Lemma 8 that $\hat{\mathcal{N}} \subset \mathcal{I}$. Then, as in the proof of Theorem 1, the boundary $\partial\hat{\mathcal{N}}$ is shown to be described only by $\mathcal{T}_n(\tilde{L})$.

Note that the resulting equation $\mathcal{T}_n(\tilde{L})$ may not be symmetric since $\tilde{L}$ is a function of a convex set $\mathcal{C}$.

Examples of $\hat{\mathcal{N}}$ compared to $\hat{\mathcal{M}}$ when the convex set constraints are a circle and a box are shown in Fig. 1.

VI. Conclusions

In this paper we studied the properties of the minimizer of the sum of strongly convex functions, in terms of the minimizers and strong convexity parameters of these functions, along with assumptions on the gradient of these functions. While identifying the region where the minimizer can lie is simple in the case of single-dimensional functions (i.e., it is given by the interval bracketed by the smallest and largest minimizers of the functions in the sum), generalizing this result to multi-dimensional functions is significantly more complicated. Thus, we established geometric properties and necessary conditions for a given point to be a minimizer. We considered two cases: one where the gradients of the functions have to be bounded by a value L at the location of the minimizer, and the other where the gradients of the functions are bounded by L everywhere inside a convex set. We used the results from the former case to provide an estimate of the region for the latter case. The boundaries of these regions are shown in Fig. 1 (in red and dark blue).

Our work in this paper focused on identifying necessary conditions for certain points to be minimizers, and thus the regions that we have characterized are overapproximations of the true regions. Future work will include finding sufficient conditions for given points to be a minimizers, and generalizing these regions to handle sums of multiple strongly convex functions.

References

[1] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Identifying suspicious URLs: an application of large-scale online learning," in *International Conference on Machine Learning*, 2009, pp. 681–688.

[2] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, 2011.

[4] A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, vol. 102, no. 4, pp. 460–497, 2014.

[5] A. Maknouninejad and Z. Qu, "Realizing unified microgrid voltage profile and loss minimization: A cooperative distributed optimization and control approach," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1621–1630, 2014.

[6] N. Li, L. Chen, and S. H. Low, "Optimal demand response based on utility maximization in power networks," in *IEEE Power and Energy Society General Meeting*, 2011, pp. 1–8.

[7] M. Schwager, "A gradient optimization approach to adaptive multi-robot control," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[8] S. Hosseini, A. Chapman, and M. Mesbahi, "Online distributed optimization via dual averaging," in *IEEE Conference on Decision and Control (CDC)*, 2013, pp. 1484–1489.

[9] E. Montijano and A. Mosteo, "Efficient multi-robot formations using distributed optimization," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 6167–6172.

[10] S. Hosseini, A. Chapman, and M. Mesbahi, "Online distributed ADMM via dual averaging," in *IEEE Conference on Decision and Control (CDC)*, 2014, pp. 904–909.

[11] M. Zhu and S. Martínez, *Distributed Optimization-Based Control of Multi-Agent Networks in Complex Environments*. Springer, 2015.

[12] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, vol. 31, no. 9, pp. 803–812, 1986.

[13] A. Nedić, A. Ozdaglar, and P. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.

[14] B. Johansson, M. Rabi, and M. Johansson, "A randomized incremental subgradient method for distributed optimization in networked systems," *SIAM Journal on Optimization*, vol. 20, no. 3, pp. 1157–1170, 2009.

[15] M. Zhu and S. Martínez, "On distributed convex optimization under inequality and equality constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2012.

[16] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *IEEE Conference on Decision and Control*, Orlando, Florida, 2011, pp. 3800–3805.

[17] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 781–786, 2014.

[18] A. Nedic and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2015.

[19] S. Sundaram and B. Gharesifard, "Secure local filtering algorithms for distributed optimization," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*. IEEE, 2016, pp. 1871–1876.

[20] L. Su and N. Vaidya, "Byzantine multi-agent optimization," *arXiv preprint arXiv:1506.04681*, 2015.

[21] K. Kuwaranancharoen and S. Sundaram, "On the location of the minimizer of the sum of strongly convex functions," *arXiv preprint arXiv:1804.07699*, 2018.